Our Docket No.: 3364P071C
Express Mail No.: EV 339918295 US

CONTINUATION APPLICATION FOR UNITED STATES PATENT

FOR

3D STEREOSCOPIC/MULTIVIEW VIDEO PROCESSING SYSTEM AND ITS METHOD

Inventor(s):
Kug-Jin Yun
Suk-Hee Cho
Yunjung Choi
Jinhwan Lee
Chieteuk Ahn

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
12400 Wilshire Boulevard, Seventh Floor
Los Angeles, California 90025
Telephone: (310) 207-3800

3D Stereoscopic/Multiview Video Processing System and its Method

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention relates to a three-dimensional (3D) video processing system and its method. More specifically, the present invention relates to an apparatus and method for processing stereoscopic/multiview three-dimensional video images based on MPEG-4 (Motion Picture Experts Group-4).

Description of the Related Art

[0002] MPEG is an information transmission method through video image compression and code representation and has been developed to the next-generation compression method, MPEG-7, subsequent to the current MPEG-1/2/4.

[0003] MPEG-4, i.e., the video streaming standard for freely storing multimedia data including video images in digital storage media on the Internet is now in common use and is applicable to a portable webcasting MPEG-4 player (PWMP), etc.

[0004] More specifically, MPEG-4 is the standard for general multimedia including still pictures, computer graphics (CG), audio coding of analytical composition systems, composite audio based on the musical instrument data interface (MIDI), and text, by adding compression coding of the existing video and audio signals.

[0005] Accordingly, the technology of synchronization among objects that are different from one another in attributes as well as the object descriptor representation method for representing the attributes of the individual objects and the scene description information representation method for representing the temporal and spatial correlations among the objects is a matter of great importance.

[0006] In the MPEG-4 system, media objects are coded and transferred in the form of an elementary stream (ES), which is characterized by variables determining a maximum transmission rate on the network, QoS (Quality of Service) factors, and necessary decoder resources. The individual media object is composed of one elementary stream of a particular coding method and is streamed through a hierarchy structure, which comprises a compression layer, a sync layer, and a delivery layer.

[0007] The MPEG-4 system packetizes the data stream output from a plurality of encoders per access unit (AU) to process objects of different attributes and freely represents the data stream using the object descriptor information and the scene description information.

[0008] However, the existing MPEG-4 system standardizes only two-dimensional (hereinafter referred to as "2D") multimedia data and therefore scarcely concerns the

technology for processing stereoscopic/multiview 3D video data.

SUMMARY OF THE INVENTION

[0009] It is therefore an object of the present invention to process stereoscopic/multiview three-dimensional video data based on the existing MPEG-4 standards.

[0010] It is another object of the present invention to minimize the overlapping header information of packets by multiplexing multi-channel field-based elementary streams having the same temporal and spatial information into a single elementary stream.

[0011] It is further another object of the present invention to select data suitable for the user's demand and the user system environments, thereby facilitating the data stream.

[0012] In one aspect of the present invention, there is provided a stereoscopic/multiview three-dimensional video processing system, which is to process video images based on MPEG-4, the system including: a compressor for processing input stereoscopic/multiview three-dimensional video data to generate field-based elementary streams of multiple channels, and outputting the multi-channel elementary streams into a single integrated elementary stream; a packetizer for receiving the elementary streams from the compressor per access unit and packetizing the received elementary streams; and a transmitter for processing the packetized stereoscopic/multiview three-dimensional video data and transferring or storing the processed video data.

[0013] The compressor includes: a three-dimensional object encoder for coding the input stereoscopic/multiview three-dimensional video data to output multi-channel field-based elementary streams; and a three-dimensional elementary stream mixer for integrating the multi-channel field-based elementary streams into a single elementary stream.

[0014] The three-dimensional object encoder outputs elementary streams in the unit of 4-channel fields including odd and even fields of a left image and odd and even fields of a right image, when the input data are three-dimensional stereoscopic video data. Alternatively, the three-dimensional object encoder outputs N×2 field-based elementary streams to the three-dimensional elementary stream mixer, when the input data are N-view multiview video data.

[0015] The three-dimensional elementary stream mixer generates a single elementary stream by selectively using a plurality of elementary streams input through multiple channels according to a display mode for stereoscopic/multiview three-dimensional video data selected by a user. The display mode is any one mode selected from a two-dimensional video display mode, a three-dimensional video field shuttering display mode for displaying three-dimensional video images by field-based shuttering, a three-dimensional stereoscopic video frame shuttering display mode for displaying three-dimensional video images by frame-based shuttering, and a multiview three-dimensional

video display mode for sequentially displaying images at a required frame rate.

[0016] The three-dimensional elementary stream mixer multiplexes 4-channel field-based elementary streams of stereoscopic three-dimensional video data output from the three-dimensional object encoder into a single-channel access unit stream using 2-channel elementary streams in the order of the odd field elementary stream of a left image and the even field elementary stream of a right image, when the display mode is the three-dimensional video field shuttering display mode.

[0017] The three-dimensional elementary stream mixer multiplexes 4-channel field-based elementary streams of stereoscopic three-dimensional video output from the three-dimensional object encoder into a single-channel access unit stream using 4-channel elementary streams in the order of the odd field elementary stream of a left image, the even field elementary stream of the left image, the odd field elementary stream of a right image, and the even field elementary stream of the right image, when the display mode is the three-dimensional video frame shuttering display mode.

[0018] The three-dimensional elementary stream mixer multiplexes 4-channel field-based elementary streams of stereoscopic three-dimensional video output from the three-dimensional object encoder into a single-channel access unit stream using 2-channel elementary streams in the order of the odd field elementary stream of a left image and the even field elementary stream of the left image, when the display mode is the two-dimensional video display mode.

[0019] The three-dimensional elementary stream mixer multiplexes N×2 field-based elementary streams of N-view video output from the three-dimensional object encoder into a single-channel access unit stream sequentially using the individual viewpoints in the order of odd field elementary streams and even field elementary streams by viewpoints, when the display mode is the three-dimensional multiview video display mode.

[0020] When processing the elementary streams into a single-channel access unit stream and sending them to the packetizer, the compressor sends the individual elementary stream to the packetizer by adding at least one of image discrimination information representing whether the elementary stream is two- or three-dimensional video data, display discrimination information representing the display mode of the stereoscopic/multiview three-dimensional video selected by a user, and viewpoint information representing the number of viewpoints of a corresponding video image that is a multiview video image.

[0021] Hence, the packetizer receives a single-channel stream from the compressor per access unit, packetizes the received single-channel stream, and then constructs a packet header based on the additional information. Preferably, the packet header includes an access unit start flag representing which byte of a packet payload is the start of the stream, an access unit end flag representing which byte of the packet payload is the end of the

stream, an image discrimination flag representing whether the elementary stream output from the compressor is two- or three-dimensional video data, a decoding time stamp flag, a composition time stamp flag, a viewpoint information flag representing the number of viewpoints of the video image, and a display discrimination flag representing the display mode.

[0022] In another aspect of the present invention, there is provided a stereoscopic/multiview three-dimensional video processing method that includes: (a) receiving three-dimensional video data, determining whether a corresponding video image is a stereoscopic or multiview video image, and processing the corresponding video data according to the determination result to generate multi-channel field-based elementary streams; (b) multiplexing the multi-channel field-based elementary streams in a display mode selected by a user to output a single-channel elementary stream; (c) packetizing the single-channel elementary stream received; and (d) processing the packetized stereoscopic/multiview three-dimensional video image and sending or storing the processed video image.

[0023] The step (a) of generating the elementary streams includes: outputting elementary streams in the unit of 4-channel fields including odd and even fields of a left three-dimensional stereoscopic image and odd and even fields of a right three-dimensional stereoscopic image, when the input data are three-dimensional stereoscopic video data; and outputting N×2 field-based elementary streams, when the input data are N-view multiview video data.

[0024] The multiplexing step (b) further includes multiplexing 4-channel field-based elementary streams of stereoscopic three-dimensional video into a single-channel access unit stream using 2-channel elementary streams in the order of the odd field elementary streams of a left image and the even field elementary streams of a right image, when the display mode is a three-dimensional video field shuttering display mode.

[0025] The multiplexing step (b) further includes multiplexing 4-channel field-based elementary streams of stereoscopic three-dimensional video into a single-channel access unit stream using 4-channel elementary streams in the order of the odd field elementary stream of a left image, the even field elementary stream of the left image, the odd field elementary stream of a right image and the even field elementary stream of the right image, when the display mode is a three-dimensional video frame shuttering display mode.

[0026] The multiplexing step (b) further includes multiplexing 4-channel field-based elementary streams of stereoscopic three-dimensional video into a single-channel access unit stream using 2-channel elementary streams in the order of the odd field elementary stream of a left image and the even field elementary stream of the left image, when the display mode is a two-dimensional video display mode.

[0027] The multiplexing step (b) further includes multiplexing N×2 field-based elementary streams of N-view video into a single-channel access unit stream sequentially using the individual viewpoints in the order of odd field elementary streams and even field elementary streams by viewpoints, when the display mode is a three-dimensional multiview video display mode.

[0028] The multiplexing step (b) includes: processing multiview three-dimensional video images to generate multi-channel elementary streams and using time information acquired from an elementary stream of one channel among the multi-channel elementary streams to acquire synchronization with elementary streams of the other viewpoints, thereby acquiring synchronization among the three-dimensional video images.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] The accompanying drawings, which are incorporated in and constitute a part of the specification, illustrate an embodiment of the invention, and, together with the description, serve to explain the principles of the invention:

[0030] FIG. 1 is a schematic of a stereoscopic/multiview 3D video processing system according to an embodiment of the present invention;

[0031] FIG. 2 is an illustration of information transmitted by ESI for the conventional 2D multimedia;

[0032] FIG. 3 is an illustration of input/output data of a stereoscopic 3D video encoder according to an embodiment of the present invention;

[0033] FIG. 4 is an illustration of input/output data of a 3D N-view video encoder according to an embodiment of the present invention;

[0034] FIG. 5 is an illustration of input/output data of a 3D ES mixer for stereoscopic video according to an embodiment of the present invention;

[0035] FIG. 6 is an illustration of input/output data of a multi-view 3D ES mixer according to an embodiment of the present invention;

[0036] FIG. 7 is a schematic of a field-based ES multiplexer for stereoscopic 3D video images for field shuttering display according to an embodiment of the present invention;

[0037] FIG. 8 is a schematic of a field-based ES multiplexer for stereoscopic 3D video images for frame shuttering display according to an embodiment of the present invention;

[0038] FIG. 9 is a schematic of a field-based ES multiplexer for stereoscopic 3D video images for 2D display according to an embodiment of the present invention;

[0039] FIG. 10 is a schematic of a field-based ES multiplexer for multiview 3D video images for 3D display according to an embodiment of the present invention;

[0040] FIG. 11 is a schematic of a field-based ES multiplexer for multiview 3D video images for 2D display according to an embodiment of the present invention;

[0041] FIG. 12 is an illustration of additional transfer information for the conventional ESI for processing stereoscopic/multiview 3D video images according to an embodiment of the present invention;

[0042] FIG. 13 is a schematic of a sync packet header for processing stereoscopic/multiview 3D video images according to an embodiment of the present invention;

[0043] FIG. 14 MPEG-4 is stream types defined by a system; and

[0044] FIG. 15 is a 3D video image stream type for processing a stereoscopic/multiview 3D video image by a decoder.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0045] In the following detailed description, only the preferred embodiment of the invention has been shown and described, simply by way of illustration of the best mode contemplated by the inventor(s) of carrying out the invention. As will be realized, the invention is capable of modification in various obvious respects, all without departing from the invention. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not restrictive.

[0046] In the embodiment of the present invention, MPEG-4 stereoscopic/multiview 3D video data are processed. Particularly, the encoded field-based elementary streams output through multiple channels at the same time are integrated into a single-channel elementary stream according to the user's system environments and the user's selected display mode, and then multiplexed into a single 3D access unit stream (hereinafter referred to as "3D_AU stream").

[0047] More particularly, the streaming is enabled to support all the four display modes: a two-dimensional video display mode, a three-dimensional video field shuttering display mode for displaying three-dimensional video images by field-based shuttering, a three-dimensional stereoscopic video frame shuttering display mode for displaying three-dimensional video images by frame-based shuttering, and a multiview three-dimensional video display mode for sequentially displaying images at a required frame rate by using a lenticula lens or the like.

[0048] To enable the multiplexing of the stereoscopic/multiview 3D video images and the above-mentioned four display defined by the user, the embodiment of the present invention generates new header information of a sync packet header and constructs the header with the overlapping information minimized. Furthermore, the embodiment of the present invention simplifies synchronization among 3D video images by using the time information acquired from one-channel elementary streams among multi-channel elementary streams for multiview video images at the same time, to acquire

synchronization with the elementary streams of the other viewpoints.

[0049] FIG. 1 is a schematic of a stereoscopic/multiview 3D video processing system (hereinafter referred to as "video processing system") according to an embodiment of the present invention.

[0050] The video processing system according to the embodiment of the present invention, which is to process stereoscopic/multiview 3D video data based on the MPEG-4 system, comprises, as shown in FIG. 1, a compression layer 10 supporting multiple encoders; a sync layer 20 receiving access unit (AU) data and generating packets suitable for synchronization; and a delivery layer 30 including a FlexMux 31 optionally given for simultaneous multiplexing of multiple streams, and a delivery multimedia integrated framework (DMIF) 32 for constructing interfaces to transport environments and storage media .

[0051] The compression layer 10 comprises various object encoders for still pictures, computer graphics (CG), audio coding of analytical composition systems, musical instrument data interface (MIDI), and text, as well as 2D video and audio.

[0052] More specifically, the compression layer 10 comprises, as shown in FIG. 1, a 3D object encoder 11, a 2D object encoder 12, a scene description stream generator 13, a object descriptor stream generator 14, and 3D elementary stream mixers (hereinafter referred to as "3D_ES mixers") 15 and 16.

[0053] The 2D object encoder 12 encodes various objects including still pictures, computer graphics (CG), audio coding of analytical composition systems, musical instrument data interface (MIDI), and text, as well as 2D video and audio. The elementary stream output from the individual encoders in the 2D object encoder 12 is output in the form of an AU stream and is transferred to the sync layer 20.

[0054] The object descriptor stream generator 14 generates an object descriptor stream for representing the attributes of multiple objects, and the scene configuration information stream generator 13 generates a scene description stream for representing the temporal and spatial correlations among the objects.

[0055] The 3D object encoder 11 and the 3D_ES mixers 15 and 16 are to process stereoscopic/multiview 3D video images while maintaining compatibility with the existing MPEG-4 system.

[0056] The 3D object encoder 11 is an object-based encoder for stereoscopic/multiview 3D video data, and comprises a plurality of 3D real image encoders for processing images actually taken by cameras or the like, and a 3D computer graphic (CG) encoder for processing computer-generated images, i.e., CG.

[0057] When the input data are stereoscopic 3D video images generated in different directions, the 3D object encoder 11 outputs elementary streams in the units of even and

odd fields of left and right images, respectively. Contrarily, when the input data are N-view 3D video images, the 3D object encoder 11 outputs N×2 field-based elementary streams to the 3D_ES mixers 15 and 16.

[0058] The 3D_ES mixers 15 and 16 process the individual elementary streams output from the 3D object encoder 11 into a single 3D_AU stream, and send the single 3D_AU stream to the sync layer 20.

[0059] The above-stated single 3D_AU stream output from the compression layer 10 is transferred to the sync layer via an elementary stream interface (ESI). The ESI is an interface connecting media data streams to the sync layer that is not prescribed by the ISO/IEC 14496-1 but is provided for easy realization, and accordingly, can be modified in case of need. The ESI transfers SL packet header information. An example of the SL packet header information transferred through the ESI in the existing MPEG-4 system is illustrated in FIG. 2. The SL packet header information is used for the sync layer 20 generating an SL packet header.

[0060] To maintain temporal synchronization between or in the elementary streams, the sync layer 20 comprises a plurality of object packetizers 21 for receiving the individual elementary stream output from the compression layer 10 per AU, dividing it into a plurality of SL packets to generate a payload of individual SL packets and to generate a header of each individual SL packet with reference to information received for every AU via the ESI, thereby completing SL packets composed of the header and the payload.

[0061] The SL packet header is used to check continuity in case of data loss and includes information related to a time stamp.

[0062] The packet stream output from the sync layer 20 is sent to the delivery layer 30, and is processed into a stream suitable for interfaces to transport environments and storage media via the DIMF 32 after being multiplexed by the FlexMux 31.

[0063] The basic processing of the sync layer 20 and the delivery layer 30 is the same as that of the existing MPEG-4 system, and will not be described in detail.

[0064] Now, a description will be given as to a method for multiplexing stereoscopic/multiview 3D video images based on the above-constructed video processing system.

[0065] As an example, 2D images and multi-channel 3D images (including still or motion pictures) taken by at least two cameras, or computer-generated 3D images, i.e., CG, are fed into the 2D object encoder 12 and the 3D object encoder 11 of the compression layer 10, respectively. The multiplexing process for 2D images is well known to those skilled in the art and will not be described in detail.

[0066] The stereoscopic/multiview 3D video images that are real images taken by cameras are input to a 3D real image encoder 111 of the 3D object encoder 11, and the CG as a

computer-generated 3D stereoscopic/multiview video image is input to a 3D CG encoder 112 of the 3D object encoder 11.

[0067] FIGS. 3 and 4 illustrate the operations of the plural 3D real image encoders and the 3D CG encoder, respectively.

[0068] When the input data are a stereoscopic 3D video image generated in the left and right directions, as shown in FIG. 3, the 3D real image encoder 111 or the 3D CG encoder 112 encodes left and right images or left and right CG data in the unit of fields to output elementary streams in the unit of 4-channel fields.

[0069] More specifically, the stereoscopic 3D real image or CG is encoded into a stereoscopic 3D elementary stream of left odd fields 3DES_LO, a stereoscopic 3D elementary stream of left even fields 3DES_LE, a stereoscopic 3D elementary stream of right odd fields 3DES_RO, and a stereoscopic 3D elementary stream of right even fields 3DES_RE.

[0070] When the input data are an N-view video image, the 3D real image encoder 111 or the 3D CG encoder 112 encodes N-view image or CG data in the unit of fields to output odd field elementary streams of first to N-th viewpoints, and even field elementary streams of first to N-th viewpoints.

[0071] More specifically, as shown in FIG. 4, the N-view video is encoded into N×2 elementary streams including an odd field elementary stream of the first viewpoint 3DES_#1 OddField, an odd field elementary stream of the second viewpoint 3DES_#2 OddField, ..., an odd field elementary stream of the N-th viewpoint 3DES_#N OddField, an even field elementary stream of the first viewpoint 3DES_#1 EvenField, an even field elementary stream of the second viewpoint 3DES_#2 EvenField, ..., and an even field elementary stream of the N-th viewpoint 3DES_#N EvenField.

[0072] As described above, the multi-channel field-based elementary streams output from the stereoscopic/multiview 3D object encoder 11 are input to the 3D_ES mixers 15 and 16 for multiplexing.

[0073] FIGS. 5 and 6 illustrate the multiplexing process of the 3D_ES mixers.

[0074] The 3D_ES mixers 15 and 16 multiplex the multi-channel field-based elementary streams into a 3D_AU stream to output a single-channel integrated stream. Here, the elementary stream data to be transferred are variable depending on the display mode. Accordingly, multiplexing is performed to transfer only the necessary elementary streams for the individual display mode.

[0075] There are four display modes: a 2D video display mode, a 3D video field shuttering display mode, a 3D video frame shuttering display mode, and a multiview 3D video display mode.

[0076] FIGS. 7 to 11 illustrate multiplexing examples for multi-channel field-based

elementary streams depending on the display mode concerned. FIGS. 7, 8, and 9 show multiplexing methods for stereoscopic 3D video data, and FIGS. 10 and 11 show multiplexing method for multiview 3D video data.

[0077] When the user selects the 3D video field shuttering display mode for stereoscopic 3D video data, the stereoscopic 3D elementary stream of left odd fields 3DES_LO and the stereoscopic 3D elementary stream of right even fields 3DES_RE among the 4-channel elementary streams output from the 3D object encoder 11 are sequentially integrated into a single-channel 3D_AU stream, as shown in FIG. 7.

[0078] When the user selects the 3D video frame shuttering display mode for stereoscopic 3D video data, the stereoscopic 3D elementary stream of left odd fields 3DES_LO, the stereoscopic 3D elementary stream of left even fields 3DES_LE, the stereoscopic 3D elementary stream of right odd fields 3DES_RO, and the stereoscopic 3D elementary stream of right even fields 3DES_RE among the 4-channel elementary streams are sequentially integrated into a single-channel 3D_AU stream, as shown in FIG. 8.

[0079] When the user selects the 2D video display mode for stereoscopic 3D video data, the stereoscopic 3D elementary stream of left odd fields 3DES_LO and the stereoscopic 3D elementary stream of left even fields 3DES_LE are sequentially integrated into a single-channel 3D_AU stream, as shown in FIG. 9.

[0080] When the user selects the 3D video display mode for multiview 3D video data, the elementary streams are integrated into a single-channel 3D_AU stream in the order of odd and even fields for every viewpoint and then in the order of viewpoints, as shown in FIG. 10. Namely, the elementary streams of a multiview video image are integrated into a single-channel 3D_AU stream in the order of the odd field elementary stream of the first viewpoint 3DES_#1 OddField, the even field elementary stream of the first viewpoint 3DES_#1 EvenField, ..., the odd field elementary stream of the N-th viewpoint 3DES_#N OddField, and the even field elementary stream of the N-th viewpoint 3DES_#N EvenField.

[0081] When the user selects the 2D video display mode for multiview 3D video data, only the odd and even field elementary streams of one viewpoint are sequentially integrated into a single-channel 3D_AU stream, as shown in FIG. 11. Accordingly, the user is enabled to display images of his/her desired viewpoint in the 2D video display mode for multiview 3D video images.

[0082] As described above, the single-channel 3D_AU stream output from the 3D_ES mixers 15 and 16 are fed into the sync layer 20. In addition to the information transferred from the ESI, as shown in FIG. 2, the single channel 3D_AU stream includes optional information for stereoscopic/multiview 3D video streaming according to the embodiment

of the present invention.

[0083] The syntax and semantics of the information added to the stereoscopic/multiview 3D video data are defined in FIG. 12.

[0084] FIG. 12 shows the syntax and semantics of the information added to the single 3D_AU stream for stereoscopic/multiview 3D video images, where only the optional information other than the information transferred via the ESI is illustrated.

[0085] More specifically, three information sets such as a display discrimination flag 2D_3DDispFlag, and a viewpoint information flag NumViewpoint are additionally given, as shown in FIG. 12.

[0086] The display discrimination flag 2D_3DDispFlag represents the display mode for stereoscopic/multiview 3D video chosen by the user. In this embodiment, the display discrimination flag is, if not specifically limited to, "00" for the 2D video display mode, "01" for the 3D video field shuttering display mode, "10" for the 3D video frame shuttering display mode, and "11" for the multiview 3D video display mode.

[0087] The viewpoint information flag NumViewpoint represents the number of viewpoints for motion pictures. Namely, the viewpoint information flag is designated as "2" for stereoscopic 3D video data that are video images of two viewpoints, and as " N " for 3D N-view video data that are video images of N viewpoints.

[0088] The sync layer 20 receives the input elementary streams per AU, divides it into a plurality of SL packets to generate a payload of the individual SL packets and constructs a sync packet header based on the information transferred via the ESI for every AU, and the above-stated additional information for stereoscopic/multiview 3D video images (i.e., display discrimination flag, and viewpoint information flag).

[0089] FIG. 13 illustrates the structure of a sync packet header that is header information added to one 3D_AU stream for stereoscopic 3D video data according to an embodiment of the present invention.

[0090] In the sync packet header shown in FIG. 13, an access unit start flag AccessUnitStartFlag represents which byte of the sync packet payload is the start of the 3D_AU stream. For example, the flag bit of "1" means that the first byte of the SL packet payload is the start of one 3D_AU stream.

[0091] An access unit end flag AccessUnitEndFlag represents which byte of the sync packet payload is the end of the 3D_AU stream. For example, the flag bit of "1" means that the last byte of the SL packet payload is the ending byte of the current 3D_AU stream.

[0092] An object clock reference (OCR) flag represents how many object clock references follow. For example, the flag bit of "1" means that one object clock reference follows.

[0093] An idle flag IdleFlag represents the output state of the 3D_AU stream. For example, the flag bit of "1" means that 3D_AU data are not output for a predetermined time, and the flag bit of "0" means that 3D_AU data are output.

[0094] A padding flag PaddingFlag represents whether or not padding is present in the SL packet. For example, the flag bit of "1" means that padding is present in the SL packet.

[0095] The padding bit PaddingBits represents a padding mode to be used for the SL packet and has a default value of "0".

[0096] A packet sequence number PacketSequenceNumber has a modulo value continuously increasing for the individual SL packet. Discontinuity in the decoder means a loss of at least one SL packet.

[0097] The object clock reference (OCR) includes an OCR time stamp and exists in the SL packet header only when the OCR flag is set.

[0098] The flag bit of the access unit start flag AccessUnitStartFlag set to "1" represents that the first byte of the SL packet payload is the start of one 3D_AU, in which case information of the optional fields is transferred.

[0099] A random access point flag RandomAccessPointFlag having a flag bit set to " 1 " represents that random access to contents is enabled.

[00100] A 3D_AU sequence number 3D_AUSequenceNumber has a module value continuously increasing for the individual 3D_AU. Discontinuity in the decoder means a loss of at least one 3D_AU.

[00101] A decoding time stamp flag DecodingTimeStampFlag represents the presence of a decoding time stamp (DTS) in the SL packet.

[00102] A composition time stamp flag CompositionTimeStampFlag represents the presence of a composition time stamp (CTS) in the SL packet.

[00103] An instant bit rate flag InstantBitRateFlag represents the presence of an instant bit rate in the SL packet.

[00104] A decoding time stamp (DTS) is a DTS present in the related SL configuration descriptor and exists only when the decoding time differs from the composition time for the 3D_AU.

[00105] A composition time stamp (CTS) is a CTS present in the related Sl configuration descriptor.

[00106] A 3D_AU length represents the byte length of the 3D_AU.

[00107] An instant bit rate represents the bit rate for the current 3D_AU, and is effective until the next instant bit rate field appears.

[00108] A degradation priority represents the priority of the SL packet payload.

[00109] A viewpoint information flag NumViewpoint represents the number of

viewpoints of motion pictures. Namely, the viewpoint information flag is set to "2" for stereoscopic 3D video data that are motion pictures of two viewpoints; or the viewpoint information flag is set to "N" for 3D N-view video data.

[00110] A display discrimination flag 2D_3DDispFlag represents the display mode for 3D video data in the same manner as the case of stereoscopic 3D video data. In this embodiment, the display discrimination flag is set to "00" for the 2D video display mode, "01" for the 3D video field shuttering display mode, "10" for the 3D video frame shuttering display mode and "11" for the multiview video display mode.

[00111] Once the above-constructed header is built, the sync layer 20 combines the header with the payload to generate an SL packet and sends the SL packet to the delivery layer 30.

[00112] After being multiplexed at the FlexMux 31, the SL packet stream transferred to the delivery layer 30 is processed into a stream suitable for an interface to transport environments via the DIMF 32 and sent to a receiver. Alternatively, the SL packet stream is processed into a stream suitable for an interface to storage media and is stored in the storage media.

[00113] The receiver decodes the processed packet stream from the video processing system to reproduce the original image.

[00114] In this case, the 3D object decoder at the receiver detects the stream format type of the multiplexed 3D_AU so as to restore the 3D video data in the stream format type of each 3D-AU multiplexed. Thus the 3D object decoder performs decoding after detecting the stream format type of the 3D_AU based on the values stored in the viewpoint information flag NumViewpoint and the display discrimination flag 2D_3DDispFlag among the information stored in the header of the packet received.

[00115] For example, when the viewpoint information flag NumViewpoint is "2" and the display discrimination flag 2D_3DDispFlag is "00" in the header of the transferred packet stream, stereoscopic 3D video data are to be displayed in the 2D video display mode and the 3D_AU is multiplexed in the order of the 3D elementary stream of left odd fields 3DES_LO and the 3D elementary stream of left even fields 3DES_LE, as shown in FIG. 10.

[00116] When the viewpoint information flag NumViewpoint is "2" and the display discrimination flag 2D_3DDispFlag is "01", stereoscopic 3D video data are to be displayed in the 3D video field shuttering display mode and the 3D_AU is multiplexed in the order of the 3D elementary stream of left odd fields 3DES_LO and the 3D elementary stream of right even fields 3DES_RE, as shown in FIG. 8.

[00117] Finally, when the viewpoint information flag NumViewpoint is "2" and the

display discrimination flag 2D_3DDispFlag is "10", stereoscopic 3D video data are to be displayed in the 3D video frame shuttering display mode and the 3D_AU is multiplexed in the order of the 3D elementary stream of left odd fields 3DES_LO, the 3D elementary stream of left even fields 3DES_LE, and the 3D elementary stream of right even fields 3DES_RE, as shown in FIG. 9.

[00118] On the other hand, when the viewpoint information flag NumViewpoint is " 2 " and the display discrimination flag 2D_3DDispFlag is "11", stereoscopic 3D video data are to be displayed in the multiview 3D video display mode, a case that cannot occur.

[00119] When the viewpoint information flag NumViewpoint is "N" and the display discrimination flag 2D_3DDispFlag is "00", multiview 3D video data are to be displayed in the 2D video display mode and the 3D_AU is multiplexed in the order of the odd field elementary stream of the first viewpoint 3DES_#1O and the even field elementary stream of the first viewpoint 3DES_#1E, as shown in FIG. 12.

[00120] When the viewpoint information flag NumViewpoint is "N" and the display discrimination flag 2D_3DDispFlag is "11", multiview 3D video data are to be displayed in the multiview 3D video display mode and the 3D_AU is multiplexed in the order of all odd field elementary streams of the first to N-th viewpoints 3DES_#1O, ..., and 3DES_#NO and all even field elementary streams of the first to N-th viewpoints 3DES_#1E, ..., and 3DES_#NE, as shown in FIG. 11.

[00121] When the viewpoint information flag NumViewpoint is "N" and the display discrimination flag 2D_3DDispFlag is "10" or "01", multiview 3D video data are to be displayed in the 3D video frame/field shuttering display mode, a case that seldom occurs.

[00122] As stated above, the receiver checks the stream format type of the 3D_AU multiplexed in the packet stream based on the values stored in the viewpoint information flag NumViewpoint and the display discrimination flag 2D_3DDispFlag of the header of the packet stream transferred from the video processing system according to the embodiment of the present invention, and then performs decoding to reproduce 3D video images.

[00123] FIG. 14 shows stream types defined by the DecoderConfigDescriptor of the MPEG-4 system, and FIG. 15 shows a new stream type for determining whether an elementary stream of the stereoscopic 3D video image output from the compression layer is 2D or 3D video image data.

[00124] While this invention has been described in connection with what is presently considered to be the most practical and preferred embodiment, it is to be understood that the invention is not limited to the disclosed embodiments, but, on the contrary, is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

[00125] As described above, the present invention enables stereoscopic/multiview 3D video processing in the existing MPEG-4 system.

[00126] Particularly, the multi-channel field-based elementary streams having the same temporal and spatial information are multiplexed into a single elementary stream, thereby minimizing the overlapping header information.

[00127] The present invention also simplifies synchronization among 3D video data by using the time information acquired from the one-channel elementary stream among the multi-channel elementary streams for multiview video data at the same time in synchronization with elementary streams of the other viewpoints.

[00128] Furthermore, the multiplexing structure and the header construction of the present invention enable the user to selectively display stereoscopic/multiview 3D video data in the 3D video field/frame shuttering display mode, the multiview 3D video display mode, or the 2D video display mode, while maintaining compatibility with the existing 2D video processing system. Hence, the present invention can perform streaming of selected data suitable for the user's demand and system environments.